International Journal of Scientific & Engineering Research Volume 11, Issue 6, June-2020
ISSN 2229-5518

Efficient breast calcifications detection using Support Vector Machine learning and Convolutional Neural Networks

Rakshitha Ravi
Department of Electrical and Computer Engineering
McGill University, Montreal, Canada
Rakshitha.ravi@mail.mcgill.ca

***Abstract* - Breast cancer is one of the most common diseases among women. Mortality due to breast cancer has been increasing for the last 50 years, making the detection of breast cancer at early stages very crucial. One of the most common symptoms of breast cancer is the presence of microcalcifications. The current diagnostic approach is BI-RADS and visual analysis by radiologists first, and then followed by biopsy if the microcalcifications are present in large numbers and unusual patterns. This existing method can be made more efficient by using machine learning algorithms like SVM(Support Vector Machine learning), reducing the role of radiologists. One of the best techniques to detect breast calcifications will be by using convolutional neural networks, as this technique has an accuracy greater than 96% .**

Index Words- Breast cancer detection, SVM breast microcalcifications detection, microcalcifications detection using ML and CNN

1. INTRODUCTION

Breast cancer starts in the cells of the breast. A cancerous (malignant) tumour is a group of cancer cells that can grow into and destroy nearby tissue. It can also spread (metastasize) to other parts of the body.[1],[2] Certain types of microcalcifications are the only indications of breast cancer, making it important that mammographers recognize which calcifications signify a benign or malignant tumour[3]

Microcalcifications(MCs) are tiny calcium deposits that appear as small bright spots in a mammogram (as illustrated in Figure 1). Individual MCs are sometimes difficult to detect, because of the surrounding breast tissue, their variation in shape, orientation, brightness and size. [4]

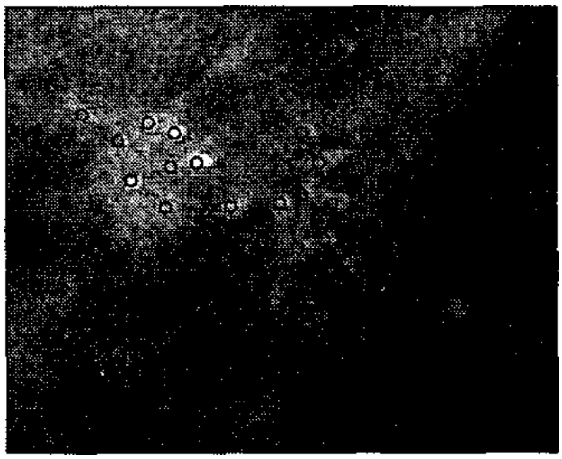

Figure 1: A section of a mammogram containing multiple MCs. [4]

The correlation between breast malignancies and breast calcifications was first made in 1913 by the German surgeon Salomon. [3]

2. BACKGROUND

. The calcium present in calcifications readily absorb the X-rays from mammograms. Calcifications typically don't show up on ultrasounds, and they never show up on breast MRIs. Calcifications are a frequent finding on mammograms, and they are especially common after menopause.

Ductal carcinoma in situ (DCIS), a type of breast cancer in which the malignancy is confined within a duct, is detected in most cases only by mammography. In a study of 100 women with DCIS, 72% of lesions presented as calcifications alone, with another 12% appearing as calcifications associated with a mass. In women younger than 49 years of age, up to 82% of DCIS manifested **as calcifications alone**. [3]

The presently used method, BIRADS (Breast Imaging Reporting and Data System), is a scheme for putting the findings from mammogram screening (for breast cancer diagnosis) into a small number of well-defined categories. If the radiologist finds the calcification to be uncommon or suspicious (a list of conditions and frequency make a calcification a suspicious one) , a biopsy is recommended to diagnose cancer.

The biggest issue to overcome is the ambiguity of the data itself, since cancer can be obscured by the natural variation of the tissue around it.[5]

A lot of research has led us to finding certain machine learning algorithms that can be used on scans to make the classification of calcification as benign or malignant, in a more efficient fashion. A few algorithms that have started gaining of attention are auto-crop algorithm , edge detection algorithm and threshold algorithm. Some of these will be explained in simple words to understand what exactly these algorithms do and how reliable they are. Support Vector Machine Learning method is a solution that uses the same physical device as BIRADS does but uses machine learning techniques to analyse the scan. Another small aspect discussed here will involve using convolutional neural networks on mammogram scans directly, and this solution also has its own benefits. [6]

One of the most important solutions, is the newly developing one. It is undergoing clinical trials but surely will be of great use in future.[7] This method is to use the new device being developed by the RF cancer detection group at McGill to detect Calcifications.[7],[8],[9],[10]

The criteria for these solutions is widely varied and includes how comfortable, painless and harmless it is to the patient and also how reliable and accurate it is. Each solution has its own advantages and its own rate of acceptance by the public. Each method has its own amounts of data to help us understand its efficiency.

3. METHODS

3.1 BIRADS

BIRADS is a system of assistance for the drafting of the reports more and more used in the world, and will soon be directly implemented on mammography and ultrasound units.[11],[12]

The evaluation of the BIRADS allows a clear synthesis of the descriptive data resulting from the use of the lexicon by the radiologist. [12]

There are 7 categories of evaluation in BI-RADS :

I. Category 0: Incomplete evaluation. Need to undergo the test again.

II. Category 1: Normal and less than 2% breast cancer risk.

III. Category 2: Benign findings present, like harmless cysts [12] .

IV. Category 3: This is the most delicate category to deal with, because of its uncertainty. [12] As a rule, for a solid mass, a follow-up check-up is recommended between 4 and 6 months later, then, every year for at least 2 years. At the end of monitoring, a stable lesion can be reclassified in BIRADS category 2.In case of change during the monitoring period, biopsy is usually performed.

V. Category 4: Suspicious and must be followed by biopsy

VI. Category 5: Highly suggestive of malignancy and further diagnosis through ultrasound or biopsy are very crucial.

VII. Category 6: Proven malignancy and further diagnosis through ultrasound or biopsy are very crucial.

The most important category for microcalcifications is category 3 and to some extent category 4. [3],[13]

With the help of morphology and distribution, calcifications can be categorized into benign, of intermediate-concern, and malignant types.[3]

Types of microcalcifications:

- REGIONAL - Calcifications are seen in a large volume, not necessarily conforming to a duct; more likely to be benign

- SCATTERED OR DIFUSED - These calcifications are seen all over the breast and they are always benign.[5],[15]

The human eye can observe the growth of the diameter and not the volume of these microcalcifications on the scans. The little nodules appear to grow less rapidly than the big ones, even if they have doubled in volume in the same time period. [12]Thus, the diameter of a nodule of 5 mm, with tumoural doubling within 6 months time, will increase only 1.25 mm making the situation more serious than diagnosed. [12]

The risky thing is that the radiologist will want to cover himself in case of a malpractice suit, and will be tempted to classify the microcalcification in the BIRADS category 4 (in the classification chart), and to have a biopsy surgery performed on it, even if it not at all required.

To avoid unnecessary surgery and pain to the patient, and also get a more accurate diagnosis, the second method can be adopted.

3.2 Support Vector Machine Learning (SVM)

The *International Symposium on Biomedical Imaging* developed the first SVM framework used to detect microcalcifications, and states it to be the best method so far.[4]

SVM learning is based on the principle of structural risk minimization [14],[15].

Instead of directly minimizing learning error, it aims to minimize the bound on the generalization error. As a result, an SVM is able to perform well when applied to data outside the training set. In recent years, SVM learning has been applied to a wide range of real-world applications where it has been found to offer superior performance to that of competing methods.

In the SVM method, a two-class pattern calcification task was performed at each location of the mammogram. The two

classes are “MC present” and “MC absent.” With an SVM formulation, a nonlinear classifier was trained using supervised learning to automatically detect the presence of microcalcifications in a mammogram.

The detailed methodology of this technique, is beyond the scope of this paper.

The proposed algorithm was compared with four other existing methods for microcalcification detection:
(1) the image difference technique (IDT)
(2) the difference of Gaussians (DoG) method
(3) the wavelet based method
(4) the two-stage multi-layer neural network method. [4].

Experimental results show that the proposed framework is quite robust over the choice of several model parameters using algorithms. In these initial results the SVM classifier outperformed all the other algorithm methods considered.

3.3 Convolutional Neural Networks on Mammograms

This method [5] utilizes a Convolutional Neural Network (CNN) to classify the calcifications and masses of different cropped images of a mammogram. The first step utilized to solve this problem is to build an image classifier to differentiate between benign and malignant, calcifications and masses found within mammograms, by classifying cropped images of the different pathologies of breast
cancer.

This method was developed based on the computing power and type of database. The dataset for this technique contained full mammograms and cropped images of the different pathology. By using the cropped images, they use less compression on the images to load them in their memory.

Finally, a sliding window image detector is constructed which pulls sections of the full mammogram into the model, one piece at a time. Each piece is passed through the classifier, predicting the possibility of the sub-image belonging to a different class of cancer.

The different sub-images
probabilities and coordinates are recorded and finally it displays the highest ranked of each sub-category of cancer.

If the known area of cancerous tissue is detected and classified properly
, the system works and there are no false positives. If the known are of cancerous tissue is detected and classifies
properly, but not displayed in the ranked area, we have a false positive elsewhere within the mammogram. [5]

If it detects the tissue but misclassify the tissue, we know the detector is working by the pathology is obscure and more analysis is needed.

Utilizing the detector and the method above, it detected 32% of cancer within the mammograms and properly classified 68% of the cancer we discovered. Only 50% of the cancer detected, or 15% overall, was detected as the most important anomaly on the mammogram. [5] This leaves a large group of false positives, with most of the false positives being outside of the breast tissue completely or within the black outside of the mammogram.
This method shows the potential of simultaneous detection and classification, and with some refinement, can be a powerful tool for diagnosing breast cancer in the future[5]

4. ANALYSIS

The first solution involves a radiologist and maybe a biopsy(depending on individual cases) to confirm. [11] The correct identification and detection of microcalcifications, using BIRADS is just

27.9%.[15],[11].The main reason for a relatively low percentage is due to human errors, and also the limit of a human .It may be the different cases, lack of concentration, even the focal capacity of our eyes, on a scan. The second solution (SVM) is more beneficial than the first one because it has an accuracy percentage of about **84.3%** as it detected 34 correctly among 38. [4]. The third technique, using convolutional neural networks is the most efficient one, as it has more than **96%** accuracy and it generates results faster and in a more detailed analysis pattern than the other two solutions. This technique, would not need a confirmation test through biopsy or any other cancer detection techniques.[15]

They later two solutions are efficient because they use algorithms and neural networks, which are proved to work more efficiently than human. They optimize results from a large amount of past data that a man would never be able to read in a lifetime. The CNN method has a better accuracy than the SVM method, mainly because Convolutional neural networks require lesser amounts of data fed, in order to give the best optimisations.[5]

ACKNOWLEDGMENTS

I would like express my heartfelt gratitude to Prof. Mark Coates and Prof. Milica Popovi´c for their patience and steadfast guidance throughout my journey at the BCD Analysis lab of the RF Breast Cancer Detection Project at McGill University from March 2019. I am grateful for the opportunity they gave me as a first year undergrad and begin my journey in the research domain. I would like to thank Prof. Steven sacks and Ms. Jane Elain Martin for insightful comments that has helped me improve this paper in innumerable ways. Lastly, I would like to express sincere appreciations to my Lab mates Lena Kranold, Collin Joseph and Chayoi Liu for useful discussions and motivating me to overcome technical challenges.

REFERENCES

[1] E. Porter, A. Santorelli, M. Coates, and M. Popović, *"Time-domain microwave breast cancer detection: extensive system testing with phantoms," Technol Cancer Res Treat.*, vol. 12, pp. 131 - 143, 2013.

[2] N Harbeck , and M Gnant . 2017. "Breast Cancer." *Lancet (London, England)* 389 (10074): 1134–50. doi:10.1016/S0140-6736(16)31891-8.

[3] LA Monda . "Differentiation of Breast Calcifications." *Radiologic Technology* 72, no. 6 (2001): 532-44.

[4] I. El-Naqa, Y Yang, M. N. Wernick, N. P. Galatsanos and R. Nishikawa, "Support vector machine learning for detection of microcalcifications in mammograms," *Proceedings IEEE International Symposium on Biomedical Imaging*, Washington, DC, USA, 2002, pp. 201-204.
doi: 10.1109/ISBI.2002.1029228

[5] G Scott, S Alex, F Ireti, and Engels, Daniel W. (2019) "The Simultaneous Detection and Classification of Mass and Calcification Leading to Breast Cancer in Mammograms," SMU Data Science Review: Vol. 2 : No. 1 , Article 10.

[6] L. Kranold, C. Quintyne, M. Coates, and M. Popović, *"Microwave Radar for Breast Screening: Initial Clinical Data with Suspicious-Lesion Patients," in Proc. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2019),* Berlin, Germany, July 23-27, 2019.

[7] E. Porter, *"Microwave breast tumor detection: simulation and design of experiments with tissue phantoms," Master's Thesis*, Department of Electrical and Computer Engineering, McGill University, Montreal, Canada, 2010.

[8] M. Coates, and M. Popović, *"Clinical Study with a Time-Domain Microwave Breast Monitor: Analysis of the System Response and Patient Attributes," in Proc. 13th European Conference on Antennas and Propagation (EUCAP 2019),* Krakow, Poland, Mar. 29 - April 5, 2019.

[9] M. Coates, and M. Popović, et al. *"Variability in Clinical Data Obtained with Flexible Time-Domain Radiofrequency Breast Monitor," in Proc. 18th International Symposium on Antenna Technology and Applied Electromagnetics (ANTEM 2018),* Waterloo, Canada, Aug. 19-22, 2018.

[10] M. Coates, and M. Popović, et.al *"Recent Advancements in Time-Domain Breast Health Screening: Observations on the Phantom Stability and Wearable Hardware," in Proc. 12th European Conference on Antennas and Propagation (EUCAP 2018),* London, UK, Apr. 9-13, 2018.

[11] P Shi, J Zhong ,A Rampun , and H Wang . "A Hierarchical Pipeline for Breast Boundary Segmentation and Calcification Detection in Mammograms." *Computers in Biology and Medicine* 96 (2018): 178-88. doi:10.1016/j.compbiomed.2018.03.011.

[12] L Levy , M Suissa , JF Chiche , G Teman , and B Martin . "Birads Ultrasonography." *European Journal of Radiology* 61, no. 2 (2007): 202-11.

[13] R Baker, K D Rogers, N Shepherd, and N Stone, 2010.New relationships between breast microcalcifications and cancer, *British Journal of Cancer* and *PMC,* 1034-1039

[14] EA Sickles . Periodic mammographic follow-up of probably benign lesions: results in 3184 consecutive cases. Journal of *Radiology* 1991; 179: 463–468.

[15] J.C. Fu, S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang, H.K. Wu, "Image segmentation feature selection and pattern classification for mammographic microcalcifications", *Computerized Medical Imaging and Graphics*, vol. 29, p. 419, 2005.